Missouri Secretary of State Robin Carnahan
Records Services Division
presents

Information Technology Essentials for Records Managers and Archivists

Workshop 1 in the Missouri Electronic Records Education and Training Initiative

March 24, 2005

Presented by:
John B. Breeden, CRM

Provided under contract with:
*e*Visory

1

Outcomes

At the end of this session, you will be able to:

- Understand the relationship of the Records Management and Information Technology disciplines
- Understand the relationship of technology and its impact on the document life cycle
- Understand Enterprise Content Management
- Be able to use information technology, records management, and document technology terms
- Understand compression, electronic document formats, database structures and networks and their impact on records management
- Be aware of some of the standards for document and records management available to assist you

2

Collaborating for Success

3

Collaborating for Success

Benefits of Collaboration

- Better storage and retrieval of information
- Greater efficiency in terms of time and money
- More consistent and accurate records
- Improved records retrieval and management
- Destruction in compliance with the law
- More flexibility, tighter security

4

Barriers

- Good records and document management software is expensive
- Records management and information technology still tend to talk different languages and have different goals in mind
- Information technology feels it is being further burdened by records and document management needs

5

And…

- Records management as a discipline was not developed to manage electronic records
- Laws to manage electronic records are still evolving
- Records management staff often are not trained in information technology
- Likewise, information technology are not trained in records management

6

Overcoming Barriers

- Providing training and education
- Supporting 'change management' by providing up to date information
- Giving independent advice about information and communication technologies

7

In Other Words

- Learn
- Communicate
- Collaborate

8

Getting Started

9

What is a record?

- **Record**: document, book, paper, photograph, map, sound recording or other material, **regardless of physical form** or characteristics, made or received pursuant to law or in connection with the transaction of official business.

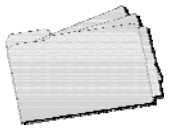
Missouri Revised Statutes, 109.210

10

What is a record?

- "Records", Information created, received, and maintained as evidence and information by an organization or person, in pursuance of legal obligations or the transaction of business.

ISO 15489



11

The Records Management Program

- The brave new world of electronic records does not negate the requirement for a sound records management program including valid and approved records retention schedules.
- If you cannot manage paper or microform, you cannot expect to manage electronic records.

12

Why Records Management?

- Records are created for business or legal reasons. Records management is all about ensuring those business records serve the organization and stakeholders by:
  - allowing government to be conducted efficiently
  - supporting management decision making
  - providing protection from disaster
  - meeting legislative and regulatory requirements
  - providing protection and support in litigation
  - maintaining organization memory

  all in the most efficient method
- A wide variety of technology is available to us to meet those objectives.

13

Elements of a Successful Records Management Program

- Authentic and reliable records
- Good records retention and disposition
- Proper access to records
- Access to records for as long as they have value

14

Elements of a Successful Records Management Program

- Archival programs
- Storage of inactive records
- Micro imaging
- Protection of Vital Records/Disaster planning

15

Special Keys to Successfully Managing Electronic Records

- Planning/Coordination in creation of electronic records
- Organization/Naming Conventions
- Metadata
- Access/Protection of Confidential records

16

Special Keys to Successfully Managing Electronic Records

- Migration
  -Hardware
  -Software
  -Media
- Disposition
- Stay tuned for Workshop 3

17

What is Needed for a Successful Electronic Records Program

- Sponsorship
- Common Goals and Vocabulary
- Collaboration, Work as a Team
- Clearly Defined Roles and Responsibilities
- Benefits for Everyone
- Great Equipment and Training
- Patience and Perseverance
- Learning from Failures and Celebrating Victories

18

Outcomes of Adequate Records Management

- Official records are created
- Official records are captured/retained
- Official records are disposed of systematically
- Access to official records is managed
- Official records can be retrieved
- Official records are trustworthy
- The management of official records is planned
- Records management training is provided to staff
- Records management reporting mechanisms are implemented
- Policies, procedures and practices exist for the management of official records

ISO 15489-2002.

19

In the Good Old Days

- Paper was king
- Microform was the "advanced storage medium"
- Distribution was physical and easy to monitor
- Destruction was tangible

20

But, Time and Technology Were on the March

- There came the computer
- And the internal network
- And finally, the external network known as

THE WEB!

21

A Strange New Land

Before computers (BC)

- Records management operated alone
- Information technology was concerned with technology, not content

After computers (AC)

- Many records managers found themselves in a strange land – the land of technology and electronic records
- Information technology found themselves trying to deal with records management and content, not just wires, boxes and coding

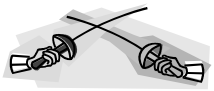
22

The Initial Results of Change

- Records management and information technology talked "different talk"
- Records and information management thought they had different missions

23

The Results, Unless



Records management and Information technology:

- Learn to communicate
- Establish common goals
- Mutually understand business processes
- Learn to work together as a team

24

Document Lifecycle

25

ECM: It's Not Your Grandparents Life Cycle Anymore!

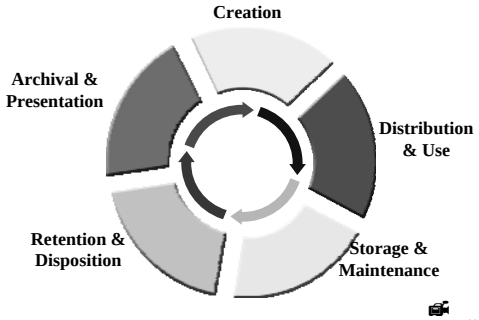
Bridging the Communications Gap

In order to bridge the communications gap, records management and information technology must learn to speak each other's language.

Brace yourselves, here come the FIRST of the definitions.

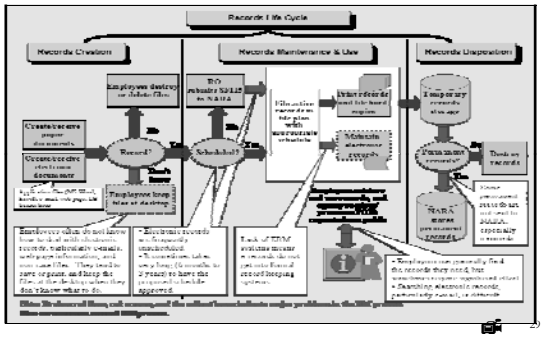
27

A Change in The Cycle of Life



28

The Brave New World



29

Document vs. Records vs. Enterprise Content Management

- **Document** – recorded information that can be treated as a unit. **ISO 15489**
- **Record-** information created, received, and maintained…in the transaction of business
- **Enterprise Content Management-** technologies, tools, and methods to preserve, and deliver content across an enterprise. **(AIIM)**

30

Records Management Terminology and Concepts

31

Definition of Record

- Any document, book, paper, photograph, map, sound recording or other material, regardless of physical form or characteristics, made or received pursuant to law or in connection with the transaction of official business (RSMo 109.210.5).

32

Record/Document Conundrum

- Records are made up of documents (Example- An Email with three attachments could be considered one record consisting of four documents
- A "document" that is not made or received pursuant to law or in the transaction of public business may not be a record but for purposes today, we are considering all documents to be records

33

Official Record

- (1) Significant, vital, or important records of continuing value to be protected, managed, and retained according to established records schedules. Often, but not necessarily an original. (2) In law, an official record has the legally recognized and judicially enforceable quality of establishing some fact.

34

Record copy

- (1) The official copy of a record that is retained for administrative, legal, fiscal, or historical purposes, sometimes the original. Duplicates of a document or multi-part form distributed to several locations may have multiple record copies, based on the purpose for which the document or form is used in each location.
- (2) The copy of a record that is captured and maintained in a recordkeeping system.

35

Records Series

- File units or documents arranged according to a filing system or kept together because they relate to a particular subject or function, result from the same activity, document a specific type of transaction, take a particular physical form, or have some other relationship arising out of their creation, receipt, or use. Also called *series*. Records schedules typically list and describe records at the record series level of aggregation.

36

Records Retention Schedule

- A listing and description of the record series maintained by all or part of an organization, prescribing the period of time that each series is to be maintained after no longer needed for current business, and when such series may be reviewed for disposition. A records schedule provides for the retention of state or local records of continuing value and for the prompt and orderly destruction of state or local records no longer possessing sufficient administrative, legal, fiscal, or historical value.

37

File Plan

- A document containing the identifying number, title or description, arrangement pattern, and disposition authority of files held in an office. A guide and aid to filing and retrieval of files.

38

Recordkeeping System

- A manual or automated system in which records are collected, organized, and categorized to facilitate their preservation, retrieval, use, and disposition.
- Any system designed to index, locate, retrieve and deliver records for use.

39

Retention Period

- The length of time a record series is to be kept after no longer needed for current business. Normally expressed either as a time period (*e.g.,* 4 yesrs), an event or action (*e.g.,* 6 months after completion of audit).

40

Disposition

- Disposition of records no longer needed for current agency business operations may include one or more of the following: reformatting, immediate destruction, immediate transfer to an archives for permanent preservation, or transfer to agency records centers prior to final disposition.

41

Archive (1)

- (1) A collection of non-current records of an organization or institution preserved because of their continuing historical value.
- (2) The organization responsible for selecting, accessioning, preserving, and making available records determined to have permanent or continuing value.
- (3) The building or portion of a building in which an archival institution is located.

42

Archive (2)

- Create a backup copy of an electronic file for non-current, but not permanent, storage.
- (RM - slang) Sometimes used inappropriately to refer to moving inactive records to off-site storage, *e.g.* "We *archived* last year's records down to the basement storeroom."

43

Questions

Break

44

Information Technology Terminology

45

Stand-Alone Computer

- Any computer, also referred to as workstation, Central Processing Unit (CPU) not connected to an internal or external (including the Web) network. The computer may, however, have alternate storage media.

46

Stand Alone Computer Terminology

- Hardware – The case which contains the physical parts of the computer, the keyboard, the monitor
  - **MOTHERBOARD** - The large circuit board into which your CPU, memory boards, and peripheral cards are plugged.
  - **HARD DRIVE** - The main device a computer uses to permanently store and retrieve information. These drives are sealed boxes typically found inside the computer. Also called a "hard disk".
  - **RANDOM ACCESS MEMORY (RAM)** - the most common computer memory which can be used by programs to perform necessary tasks while the computer is on; an integrated circuit memory chip allows information to be stored or accessed in any order and all storage locations are equally accessible .
  - **READ ONLY MEMORY (ROM)** - This is a computer's unchangeable memory. It's used to store programs that start the computer and run diagnostic functions.

47

Computer Software

- **Operating system** - the system software that controls the computer.
- **Application** (1) Software designed to perform a particular task: word processing or spreadsheet, for example. (2) A work process accomplished by a combination of various application software programs, such as using word processing, data base, and spreadsheet programs to merge address and statistical chart data into letters to be mailed to customers.

48

Code

- Computer software is written in a code. The smallest unit within the code is called the:
  - **bit** - the smallest piece of information used by the computer. Derived from "binary digit". In computer language, either a one (1) or a zero (0).

49

Computer Software Coding Terms

- **ASCII -** Pronounced "ass-kee." This is a standard means of representing characters, consisting of 256 characters. The first 128 characters are standardized, and the first 32 of those are control codes, which don't really represent visible characters but rather codes that can be used for text formatting or actions, such as making the computer beep or clearing the screen. After the 32 control codes, the next 96 standardized characters represent numbers, letters (both uppercase and lowercase), and standard punctuation marks. The last 128 characters represent different things on different platforms. ASCII is being largely supplanted by Unicode.

50

Computer Software Coding Terms

- **byte** - a piece of computer information made up of eight bits that represents a number, letter, etc.
- **driver** - a file on a computer which tells it how to communicate with an add-on piece of equipment (like a printer).

51

Storage Media

52

Storage Media

Normally, the storage media for a stand-alone computer includes the hard drive and the Compact Disk/Digital Versatile Disk (CD/DVD) drive

53

Compact Disk (CD)

- A small optical disk on which text, data, sounds, visual images, and the like can be recorded digitally and then read by a laser beam, decoded, and transmitted to a computer, television, or playback device. Data is written and read by laser beams, and is randomly accessible.

54

Digital Versatile Disc (DVD)

- 120mm optical disc on which digital video, audio, data, and images can be stored. Available in read-only, recordable, and rewritable formats.

55

CD/DVD Terms

- ROM – Read Only Memory. This disk can be created and read, but files on the disk cannot be changed or saved as they are. They must be copied and saved on a different storage medium. This does not mean you cannot add additional files to the disc once it is created.
- READ/WRITE (RW) – Files can be stored, created, and changed.

56

Document Management

57

Document Management Technologies

- Office Automation including word processing and spreadsheets
- Email management systems
- Imaging systems
- Optical Character Recognition (OCR)/Forms Recognition
- Workflow

58

Why Document Management Technologies?

- You must find information before you can use it. All of the information in the world is worthless if you can't find it. Finding information and protecting it from unauthorized use is where document management applications come into play.
- The various systems we use to index, find and organize discrete collections of information are document management systems.
- Good document management systems can improve decision making, improve customer service and save money

59

Document Management Lifecycle

- Creation/Capture
- Manage
- Storage/Retrieval
- Revision/Collaboration
- Distribution/Publication
- Preservation/Disposal

60

Some Planning Considerations for Document Management Systems

- Identify purpose of system
- Scalability
- Legality Issues
- Technology Issues
- Indexing
- Cost/Benefit Analysis

61

Document Management Terminology

62

Document Management Terminology

- ***Access tied to user ID:*** **The product limits access to system objects based on the user identification.**
- ***Approximate or pattern match (fuzzy logic):*** **Allows imprecise or ambiguous search criteria to provide an exact hit; for example, a search for "fuzzy" would count "fuzy" as a hit.**
- ***Automatic index creation:*** **The product creates indexes to each document based on the contents of the document.**
- ***Boolean expressions:*** **Search for strings that meet requirements defined by Boolean operators, e.g., find all > d.**

63

Document Management Terminology

- ***Check-in/check-out:*** **Authority to review or revise document.**
- ***Conflict detection:*** **Identifying conflicts that occur during the version merging process; these conflicts may need to be resolved manually.**
- ***Content-based retrieval:*** **Retrieval based on the content of the target document rather than based on meta data associated with the target document.**

64

Document Management Terminology

- ***Entity attributes and history:*** **Dates of creation/modification, who made modifications, what modifications were made, etc., are maintained and associated with objects.**
- ***Full-text search:*** **Searching for all occurrences of a user-specified string in the document.**

65

Document Management Terminology

- **Hierarchical storage management (HSM) (IT) A data storage management strategy in which special software is used to separate actively-used and inactive computer data by moving files between primary (on-line), secondary (near-line), and sometimes tertiary (off-line) storage media.**
- ***Hypertext links:*** **Links that tie logically associated information in a non-sequential fashion.**

66

Document Management Terminology

- ***Manage graphics and text:*** **Allows graphics and text to be stored, retrieved, associated, and tracked.**
- ***Manage nonelectronic data:*** **Pointers are provided to non-electronic data and that pointer is treated as a desktop object; controls access to it, ties attributes to it.**
- ***Multitasking/windowing:*** **Display of multiple windows in a fashion that appears to be simultaneous to the user.**

67

Document Management Terminology

- ***Nonproprietary database:*** **The database is compatible with third-party products that may be used to further automate an organization's process.**
- ***Notification of changes:*** **Based on the team definition for a document, the tool notifies team members of the changes to the objects of interest to them.**
- ***OCR conversion:*** **Conversion of raster image text to eight bit coded (ASCII) text.**

68

Document Management Terminology

- **Page-by-page viewing: The product allows the user to view the document a page at a time.**
- ***Parallel development paths:*** **Maintains version and revision control of two versions of a single document.**
- ***Password protection:*** **Access to the objects under the product's control is restricted based on a password string input by the user.**
- ***Privilege levels:*** **Ability to read, write, and delete are associated with user ID and team definition for a given object.**

69

Document Management Terminology

- ***Proximity search:*** **Allows user to specify two words that must appear within a desired number of words of each other.**
- ***Query by example:*** **The user specifies fields to be displayed, intertable linkages, and retrieval criteria directly onto forms displayed on the screen. These forms are a direct pictorial representation of the table and row structures that make up the database. Thus, the construction of a query becomes a simple check-off procedure from the viewpoint of the user.**

70

Document Management Terminology

- ***Signature block or electronic signature:*** **The product allows user to pass the document to the next phase or to indicate acceptance of the document.**
- ***Soundex matches:*** **A "hit" is defined phonetically rather than by character patterns. For example, photo would be a hit if the query entered was foto.**
- ***SQL support:*** **The product uses one of several commercially available database engines that uses ANSI SQL rather than a database with a closed architecture.**

71

Document Management Terminology

- ***Stage management:*** **Controls the movement of the document from phase to phase; verifying that the appropriate people have signed off. Access to a document is based on the document's phase.**
- ***Sticky note commenting:*** **Allows reviews to attach comments to the document.**
- ***Thesaurus:*** **Maps data/terminology links for more expansive search alternatives.**
- ***True WYSIWYG:*** **What you see on the display screen is what is printed. This must be true for graphics, tables, and equations, as well as straight text.**

72

Document Management Terminology

- Truncation: Allows the user to search for a character string regardless of the remainder of the characters that also constitute the word in which the character string is included.
- User-defined query: The product allows the user to specify a query and save it for repeated invocations.
- *Version control:* Maintaining multiple versions of a document; supporting the efficient and convenient archiving and retrieval of multiple versions; proving an audit trail of when and why changes were made.
- *Version merging:* The product merges two versions of a document.

73

Document Management Terminology

- *Weighted search:* Allows the user to assign a relevance weight to search terms.
- *Wild card:* Search for the specified string without regard for characters adjacent to the string. The syntax for a command line version of this search often uses * to represent one or more characters of any value.
- *Workflow management:* The products allows document review, path definition, reviewer comments, signature block, etc.

74

Table: Comparison of Storage Media

	Cost of Drive	Capacity (GB)	Transfer Rate (MB/sec)	Unit Media Cost	Cost/GB for media	# Needed for 20 GB
CD-R	$200	0.65	1.2 (8x unit)	$1.50	$2.31	31
CD-R/W	$250	0.65	1.2 (8x unit)	$4	$6.15	31
DVD-R (4.7GB)	$5,400	4.7	1.4	$45	$9.57	4
DVD-R (3.95 GB)	$5,400	3.95	1.4	$40	$10.13	5
DVD-RAM (single)	$500	2.6	1.4	$30	$11.54	8
DVD-RAM (double)	$500	5.2	1.4	$40	$7.69	4
DLT-IV (Digital Linear Tape)	$2,000+	20	1.5	$70	$3.50	1
Magnetic Disk	$160-$1,000	20	5–40	*	$8 - $50	1
Jaz	$300	2	4.9–8.7	$100	$50	10
Zip	$150	0.25	.8–2.4	$15	$60	80

* for magnetic disks, the cost is based on the drive itself; there are no separate media costs. Magnetic disks have substantial variance in pricing reflecting interface (IDE or SCSI), speed, and reliability. Courtesy of Tim Au Yeung, Manager of Digital Initiatives, Information Resources Press, University of Calgary.

75

Compression

76

Data Compression

- Storing data in a format that requires less space than usual.
- Data compression is particularly useful in communications because it enables devices to transmit or store the same amount of data in fewer bits. There are a variety of data compression techniques, but only a few have been standardized. The CCITT has defined a standard data compression technique for transmitting faxes (Group 3 standard) and a compression standard for data communications through modem (CCITT V.42*bis*). In addition, there are file compression formats, such as ZIP.
- Data compression is also widely used in backup utilities and database management systems.

77

Data Compression Choices

- Lossy: Data compression technology that attempts to eliminate redundant or unnecessary information. Used for JPEG and MPEG files.
- Lossless Data compression technology that records every pixel.
- Issues include space/cost and ease of transmission considerations versus quality of document and authenticity for legal purposes.

78

Electronic Documents

79

Electronic Document Requirements

- Whatever strategy is adopted, the document management system must:
  - provide adequate context information for documents;
  - provide means to prove the authenticity of documents used as evidence;
  - provide for the ultimate disposal or preservation of records;
  - be robust against organizational or technological change;
  - provide levels of support for different types of document that accord with agency policy; and
  - provide links between paper and electronic documents.

80

Problems with Electronic Documents

- Confusion between different versions of a document (e.g. because there may be multiple copies, none of which is the authoritative version);
- Loss or destruction of documents that should be kept (e.g. because there is no central repository analogous to the paper file repository, and the author is unaware of the need for retention);
- Questionable authenticity because of possible manipulation of text in electronic documents;
- loss of context of documents (e.g. because related documents are not linked or kept together);

81

Problems with Electronic Documents

- Document context In electronic systems, documents are stored as discrete entities, without any necessary relationship to other documents. In a business environment, documents rarely occur in isolation. They may, for example, be part of a transaction, part of a discussion on a topic, or a progress report on a project. They may refer to other documents. These relationships are part of a document's context. The context is important in locating and retrieving documents and groups of documents.

82

Problems with Electronic Documents

- Authenticity How do we know that a retrieved electronic document is a correct representation of the original document? If we wish to use it as evidence, how do we prove that it is what it purports to be?

83

Problems with Electronic Documents

- Disposal of documents and records Disposal of documents and records is dictated by the Missouri State Records Commission. How do we do that with electronic documents?

84

Problems with Electronic Documents

- **Robustness against organizational change**
  Government agencies are subject to internal reorganization, splitting into multiple agencies, mergers with other agencies, and transfer of functions to and from other agencies, on a timescale which is short compared with that required for records management.

85

Problems with Electronic Documents

Robustness against technological change

- **Electronic documents rapidly become unreadable due to changes in hardware, software, and media.**
- **Electronic documents becoming inaccessible because of technological change (e.g. changes in software or storage media make the files unreadable).**

86

Problems with Electronic Documents

Links to paper systems

- **Where paper and electronic documents exist within the same agency, links between documents in the two media must be possible.**

87

Common Electronic Document Formats

- .tiff
- .jpeg
- .doc
- .wpd
- .rtf
- .pdf
- .txt
- .html

88

Questions

Lunch

89

Databases

90

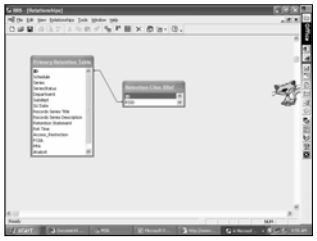
How Databases Work

- In a typical transaction processing system, business records are not stored as stable, finite, physical entities. Rather, these systems create records by combining and reusing data stored in discrete units organized into tables. Once created, a record of a business process may not, indeed, likely will not be captured as a physical entity. Not only will the record not be captured at the time of creation, it may be impossible to recreate at some later date.

91

Simple Database Structure – Tables and Fields are Linked Together

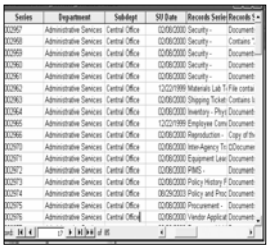
- Fields of individual information are placed together in tables
- Tables may be linked to one another
- The database will allow the user to draw information from either a single table or various fields from linked tables



92

Records

- **A group of fields that describe a particular individual, function, etc. are called a record.**



93

Records As Forms

- For ease of entry and use, records are often displayed as forms
- Forms can draw information from one or more tables

94

Problems with Database Records

- Databases are dynamic, volatile systems, in a state of continual change. Data updates occur frequently, and with DBMS software managing the system, these revisions are made in every file containing that revised data element. Moreover, databases typically maintain only the current value for any given data element. As a result, in a typical transaction processing system, it can be difficult to determine how the database looked at a given point in time.

95

Challenges

- For archivists and records managers this new architecture presents many new and difficult challenges for capturing, accessing and describing records. With the emergence of database views and dynamic and virtual documents, the differences in the way paper and electronic records are created and managed is accentuated. The widespread use of personal computers and portable devices has an equally destabilizing effect on the management of records. By creating a less structured, less centralized environment for record creation and use, in which records are frequently not integrated into the normal business processes, actions, archival and records management principles and practices needed to be reviewed and perhaps revised.

96

The Questions We Need to Answer Are…

- What is a record in an automated environment?
- How will archivists identify and appraise records?
- What documentation must be present to create a reliable and authentic record?
- What is a recordkeeping system in an automated environment? How will the system manage these records?

97

The Questions We Need to Answer Are…

- How will archivists and records managers preserve authentic electronic records for as long as is necessary? How do we keep records alive in an automated environment?
- How will access and physical custody of electronic records be managed?
- What is the overall role of the archivist/records manager in the information system development process and in the overall information technology environment?

98

More Challenges and Issues

- The primary challenge is associated with the basic but extremely important recognition that unlike paper documents, electronic records are logically constructed and often "virtual" entities. Consequently, electronic documents cannot be viewed in the same way as paper records, where so much of the content, context and structural metadata is embedded in or is part of the record. In automated systems, the vital metadata, if it exists at all, may or may not be physically associated with the content data. Vital links between metadata and the record content data may exist only in computer software programs. In some cases, the metadata may actually not be a part of the automated system at all, but may exist only as a paper document totally disassociated with the records it is describing.

99

Database Basics

- **Desktop Databases** You're probably familiar with at least one desktop database product. The market is dominated by brand-names like Microsoft Access, FileMaker Pro and Lotus. These products are relatively inexpensive and are great for single-user or non-interactive web applications.
- **Server Databases** If you're planning a heavy-duty database application like an e-commerce site or a multi-user database, you're going to need to call on one of the big guns. Server databases like Microsoft SQL Server and Oracle provide real firepower but carry a correspondingly heavy price tag.

100

Types of Databases

- There are many different types of databases, including:
  - Flat-file text databases
  - Relational databases
  - Network databases
  - Hierarchical databases such as LDAP

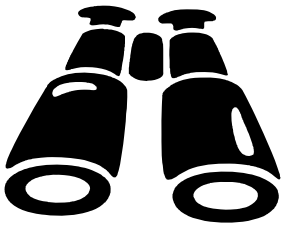
101

Flat Files and Relational Databases

- Flat files are data files that contain records with no structured relationships. In order to change the way you access information, the tables in the database must be reorganized.
- A relational database is a collection of data items organized as a set of formally-described tables from which data can be accessed or reassembled in many different ways without having to reorganize the database tables.

102

A Brief Look Inside a Database



103

Database Components

- Databases are designed to offer an organized mechanism for storing, managing and retrieving information. They do so through the use of tables. If you're familiar with spreadsheets like Microsoft Excel, you're probably already accustomed to storing data in tabular form. It's not much of a stretch to make the leap from spreadsheets to databases. Let's take a look.

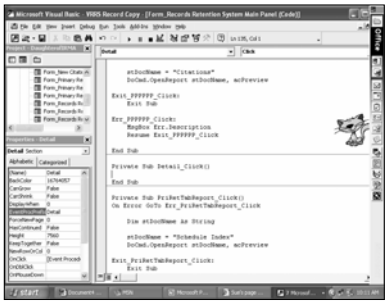
104

Database Components

- Just like Excel tables, database tables consist of columns and rows. Each column contains a different type of attribute and each row corresponds to a single record. For example, imagine that we were building a database table that contained names and telephone numbers. We'd probably set up columns named "FirstName", "LastName" and "TelephoneNumber." Then we'd simply start adding rows underneath those columns that contained the data we're planning to store.
- If we were building a table of contact information for our agency that has 50 employees, we'd wind up with a table that contains 50 rows.

105

Database Languages



106

Database Language

- **What languages are used to interact with databases?**

SQL (Structured Query Language)

By far the most common language used to interact with relational databases. Originally developed for use with IBM's DB2, the standard -- often pronounced "sequel" -- is promoted in various formats by both the American National Standards Institute (ANSI) and the International Standards Organization (ISO).

107

Database Languages

- **C**

  C is a popular language used in professional and commercial programs. Many of the major application programs and operating systems are written in C. It is a compiled, procedural language that provides both high-level commands and low level access to hardware.

  C++

  C++ is an object-oriented version of C.

108

Database Languages

- **Visual Basic**
  VB is an example of an event driven language. An event driven language helps programmers easily create programs that must constantly check for and respond to a set of events, such as key presses or mouse actions. Most GUI programs are event driven - they display controls such as menus on the screen and take action when the user activates one of the controls. When creating an event-driven program, graphical objects (*buttons, dialogue boxes, scroll bars, icons etc.*) are selected and code segments are embedded into the control. Although other languages can be used to write windows operating systems an event driven language simplifies the development process.

109

Database Terminology

110

Database

- A set of data, consisting of at least one file or of a group of integrated files, usually stored in one location and made available to several users at the same time for various applications. Databases are organized into files called tables. These tables provide a systematic way of accessing, managing, and updating data. A relational database is one that contains multiple tables of data that relate to each other through special key fields. Relational databases are far more flexible (though harder to design and maintain) than what are known as flat file databases, which contain a single table of data.

111

Database Terminology

- **Flat-File Databases:** Data is stored in files consisting of one or more readable files, usually in text format.
- **Hierarchical Databases:** Data is stored in tables with *parent/child* relationships with a strictly hierarchical structure.
- **Network Databases:** Similar to the hierarchical model, but allows more flexibility; for example, a child table can be related to more than one parent table.
- **Object-Oriented Databases:** The object-oriented database model was developed in the late 1980s and early 1990s to deal with types of data that the relational model was not well-suited for. Medical and multimedia data, for example, required a more flexible system for data representation and manipulation.
- **Object-Relational Databases:** A hybrid model, combining features of the relational and object-oriented models.

112

Database Terminology

- ActiveX Data Objects (AD0) - an API from Microsoft that lets a programmer writing Windows applications access a database.
- autonomous transaction - in Oracle's database products, an independent transaction that is initiated by another transaction.
- block - in some databases, the smallest amount of data that a program can input or output.
- catalog - a directory of information about data sets, files, or a database.

113

Database Terminology

- data dictionary - a collection of descriptions of the data objects or items in a data model for the benefit of programmers and others who need to refer to them.
- data modeling - the analysis of data objects and the identification of the relationships among these data objects.
- data mart - a repository of data gathered from operational

114

Database Terminology

- DAO (Data Access Objects) - API available with Visual Basic that lets a programmer request access to an Access database.
- data - information that has been translated into a form that is more convenient to move or process.
- data availability - assurance that data continues to be available at a required level of performance in any situation.

115

Database Terminology

- data set - in an IBM mainframe operating system, a named collection of data that contains individual data units formatted in a specific way.
- data source name - a data structure that contains the information about a specific database that an ODBC driver needs in order to connect to it.
- data warehouse - a central repository for all or significant parts of the data that an enterprise's various business systems collect.

116

Database Terminology

- DBMS - a program that lets one or more computer users create, manage, and access data in a database.
- DB2 - a family of relational database management system (RDBMS) products from IBM that serve a number of different operating system platforms.
- DDBMS (distributed database management system) - a centralized application that manages a distributed database as if it were all stored on the same computer.
- Metadata Data about the data; the description of the data resources, its characteristics, location, usage, and so on. Metadata is used to identify, describe, and define user data.

117

Database Terminology

- ODBC (Open Database Connectivity) - open standard API for accessing a database.
- operational data store (ODS) - a type of database often used as an interim area for a data warehouse.
- primary key - a key in a relational database that is unique for each record, used to identify a particular record.
- quad tree - a method of placing and locating files in a database.

118

Database Terminology

- query - a request to a database for information, or to update, modify, or delete information.
- schema - the organization or structure for a database.
- splay tree - a self-adjusting search algorithm for placing and locating files in a database.
- SQL (Structured Query Language) - a standard interactive and programming language for getting information from and updating a database.

119

Database Terminology

- table - in a relational database, a data structure that organizes the information about a single topic into rows and columns.
- tree structure - an algorithm for placing and locating records in a database.
- view - a temporary organization of data drawn from one or more tables to enable a specific perspective.

120

Web Databases

- **Web-enabled Databases** Nowadays, almost every database application calls for some kind of web interaction. Many people assume that if you're looking to publish your database on the Internet, you need to use a server database. That's not necessarily true -- a desktop database could (inexpensively!) meet your needs.

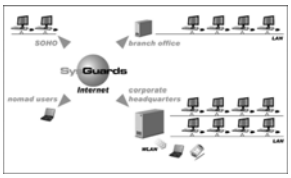
121

Network Basics

122

Networks

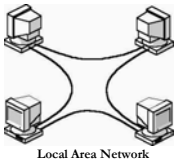
- The primary reasons for networking computers are to share information, to share hardware and software, and to centralize administration and support.



123

Networked Computers

- Networked computers are hardware devices connected by cable or wireless transmission to either each other or a central computer or computers called servers.



Local Area Network (LAN)

Wide Area Network (WAN) (The Web is a WAN)

124

Network Types

- A *local area network* (***LAN***) connects computers and resources in a limited geographical area, such as a floor, a building, a cluster of buildings, or a city. A *wide area network* (***WAN***) connects two or more local area networks through high-speed data communication lines, or connects computers and resources located more than one mile apart.

125

Network Design

- The two types of high-level network design are called client-server and peer-to-peer.
- Client-server networks feature centralized computers that store email, Web pages, files and or applications. On a peer-to-peer network, conversely, all computers tend to support the same functions. Client-server networks are much more common in business and peer-to-peer networks much more common in homes.

126

More About Network Types

- Peer-to-Peer, this is commonly two PC connected together to share files or a printer.
- Local Area Network (LAN), this is probably the most popular in smaller applications. This comes down to many computers interconnected in an office or home. This can be anywhere from two computers or hundreds in one physical area.

127

More About Network Types

- Wide Area Network (WAN), boils down to computer networks that are far apart and connect via phone, satellite, and so forth.
- Node, this is just another name for the client computer or the computer using the network. You are in essence a node right now using this website which is located on a server over the Internet. Bet you didn't node that did you? Ok, I know that was lame but what the hey!
- Topology, is the geometric arrangement of the network. This is more or less how the network is arranged. We will discuss this one in more detail.

128

Network Basics

- The Basic Network is made up of the following:
  - Server or client workstation and peripherals
  - Networking Interface Card's (NIC)
  - Cabling, hubs, and routers
  - Networking Operating System such as Windows NT/2000

129

Network Components

- **Server** - a central computer dedicated to sending and receiving data from other computers over a computer network, *e.g.*, file server, print server, email server, etc.
- **Peripherals** - an add-on component to your computer. which can communicate with a computer to:
  - Enter data, e.g. keyboards, scanners, etc.
  - Retrieve data, e.g. printers, monitors, etc.

130

Network Components

- **protocol**: a set of communication rules to make sure that everyone speaks the same language
- **cable**: the medium to connect all of the computers together
- **Network Interface Card:** Also known as an Ethernet card, or NIC card. The NIC is probably the most common component of any network. It is the component that provides the connection between a computers internal bus and the network media or cabling.

131

Networking Components

- **Hub:** A physical device that serves as a central connection point for several network devices. At its simplest level a hub is nothing more than a multi-port repeater. The hub will repeat what it receives on one port to all its other ports.
- **Router:** Usually a physical device that connects two networks and allows data to be sent and received between them. A router will determine the best path for the data to take from its source to its destination.

132

Network Topology

- A network topology represents its layout or structure from the point of view of data flow. In so-called "bus" networks, for example, all of the computers share and communicate across one common conduit, whereas in a "star" network, all data flows through one centralized device.

133

How Networks Talk to One Another

- In networking, the communication language used by computer devices is called the protocol. Yet another way to classify computer networks is by the set of protocols they support. Networks often multiple protocols to support specific applications. Popular protocols include TCP/IP, the most common protocol found on the Internet and in home networks.

134

Network Communications

- Protocol, the protocol is a ways of communication for the network. Think of it as trying to speak French to an English computer. We need a protocol so they can understand one another.
- Data Packets, messages of data that has been sent over your network.

135

Networking Protocols

- **Network Protocols:** Computers and devices on a network have to communicate with each other. A protocol is a predefined set of rules that dictate how the computers and devices communicate and exchange data over a network.
- **TCP/IP:** This is a type of network protocol. It combines the Transmission Control Protocol (TCP) and the Internet Protocol (IP). This is the protocol most commonly used in Windows based networks. It is also the standard protocol used for communications over the Internet. It was originally developed by the Department of Defense.
- **IP Address:** This is an address used by the Internet Protocol to identify a computer or device's location on the network. This number is usually assigned to a computer by the Internet Service Provider (ISP) or by the network administrator.

136

Internet!!

- A worldwide network of computers that allows public access to send, store, and receive electronic information over public networks. It is a network of networks.

137

Web Sites

- A collection of web pages stored upon and delivered by a particular web server. A web server is a computer (or set of computers) connected to the Internet that runs software that awaits and responds to requests from browsers.

138

Web Portals

- A Web site or service that offers a broad array of resources and services, such as e-mail, forums, search engines, and on-line shopping malls. The first Web portals were on-line services, such as AOL, that provided access to the Web, but by now most of the traditional search engines have transformed themselves into Web portals to attract and keep a larger audience.

139

Questions

Break

140

Enterprise Content Management (ECM)

141

Enterprise Content Management

- **Enterprise Content Management** is the technologies, tools, and methods used to capture, manage, store, preserve, and deliver content across an enterprise. At the most basic level, ECM tools and strategies allow the management of an organization's unstructured information, wherever that information exists. Numerous terms are used, depending on whom you're talking to, nearly interchangeably with ECM-integrated document management, digital asset management, integrated document and content management, and total content management to name a few. **(AIIM)**

142

Enterprise Content Management

- Regardless of the precise terminology, ECM capabilities manage traditional content types (images, office documents, graphics, drawings, and print streams) as well as the new electronic objects (Web pages and content, email, video, and rich media assets) throughout the lifecycle of that content.

143

Enterprise Content Management (ECM) Terminology

- Capture
- Manage
- Store
- Collaborate
- Delivery
- Preserve

144

(ECM) Terminology: Capture

- **Cold/ERM**-Computer Output to Laser Disk/Enterprise Report Management)-Stores and indexes computer output (reports primarily) on magnetic disks, optical discs, and magnetic tape. Once stored, the reports can be retrieved, viewed, printed, faxed, or distributed to the Internet.
- **Document Imaging** -Process of capturing, storing, and retrieving documents regardless of original format, using micrographics and/or electronic imaging (scanning, OCR, ICR, etc.).
- **E-Forms/Web Forms**-Forms designed, managed, and processed completely in an electronic environment.

145

(ECM) Terminology: Capture

- **Forms Processing**-The ability for software to accept scanned forms and extract data from the boxes and lines to populate databases. Software usually includes the ability to drop out the form so that recognition accuracy improves. ICR automatically identifies document types from the layout and structure of the document.
- **ICR (Intelligent Character Recognition)**-Advanced form of OCR technology that may include capabilities such as learning fonts during processing or using context to strengthen probabilities of correct recognition or that can recognize handprint characters.
- **Indexing** -Identification of specific attributes of a document or database record to facilitate retrieval.

146

(ECM) Terminology: Capture

- **OCR (Optical Character Recognition)**-Technique by which images of characters can be machine-identified, then converted into computer processable codes.
- **OMR (Optical Mark Recognition)** -Detects presence, or absence, of marks in defined areas; used for processing questionnaires, standardized tests, etc.
- ***XML (*eXtensible Markup Language)** -An established standard, based on the Standard Generalized Markup Language, designed to facilitate document construction from standard data items. Also used as a generic data exchange mechanism.

147

(ECM) Terminology: Manage

- **Collaboration**-Tools (collaborative authoring, video conferencing, shared whiteboards, etc.) that allow multiple users to work on the same content in a common environment.
- **Document Management**-Software that controls and organizes documents throughout an enterprise. Incorporates document and content capture, workflow, document repositories, COLD/ERM and output systems, and information retrieval systems.
- **Web Content Management** -A technology that addresses the content creation, review, approval, and publishing processes of Web-based content.

148

(ECM) Terminology: Manage

- **Records Management** - Information created, received, and maintained as evidence and information by an organization or person, in pursuance of legal obligations or the transaction of business.

  ISO 15489
- **Workflow/BPM (Business Process Management)** - Automation of business processes, in whole or in part, where documents, information, or tasks are passed from one participant to another for action, according to a set of rules. A business process is a logically related set of workflows, worksteps, and tasks that provide a product or service to customers. BPM is a mix of Process Management/Workflow with Application Integration technology.

149

(ECM) Terminology: Store

- **CD-ROM** *(Compact Disc Read Only Memory)* - Optical disc that is created by a mastering process and used for distributing read-only information.
- **Content Management System** - The capability to manage and track the location of, and relationships among, content within a repository.
- **Data Warehouse** - Central repository for all, or most, of an organization's structured data.

150

(ECM) Terminology: Store

- **DVD** *(Digital Versatile Disc)* - 120mm optical disc on which digital video, audio, data, and images can be stored. Available in read-only, recordable, and rewritable formats.
- **File System** - The way in which files are named and where they are placed logically for storage and retrieval, most commonly in a hierarchical (tree) structure.
- **Magneto Optical** *(MO)* - Recording data using a combination of magnetic and optical means to change the polarity of a magnetic field in the recording medium. Data is erasable and/or rewritable.

151

ECM Terminology: Store

- **Magnetic Storage** - Hard disks on down to floppies.
- **NAS** ***(Network Attached Storage)***-Can be part of a SAN. Hard disk storage directly attached to the network to provide information access.
- **Optical Disc** - Medium that will accept and retain information in the form of marks or density modulation in a recording layer that can be read with an optical beam.
- **RAID** ***(Redundant Array of Independent Disks)*** - Storing the same data on multiple hard disks for improved performance and fault tolerance.

152

Enterprise Content Management Terminology

- **Repositories** - Part of a Document Management system; specific functionality to control the check-in/out of material, version control, and look-up against defined attributes.
- **SAN** ***(Storage Area Network)*** - A high-speed network that connects computer systems and storage elements and allows movement of data between computer systems and storage elements and among storage elements.

153

ECM Terminology: Store

- **Tape** -A magnetic storage media. Standard widths are 8mm, 1/8-inch, 1/4-inch, 1/2-inch, 4mm DAT (Digital Audio Tape), and DLT (Digital Linear Tape) in either rolls or cassettes.

154

ECM Terminology: Delivery

- ***COLD/ERM*** - Computer Output to Laser Disc/Electronic Report Management.
- **Compression** - Technique used to reduce the number of bits in a digital image file; JPEG and TIFF are two examples.
- **Digital Rights Management** - Enables secure distribution, and disables illegal distribution, of paid content over the Web.
- **Digital Signature** - Electronic signature that can be used to authenticate the sender of a message.

155

ECM Terminology: Delivery

- **PDF** (Portable Document Format) - Format developed by Adobe Systems for document publication.
- **Personalization** - Matching content to the individual.
- **PKI** (Public Key Infrastructure) - Enables the secure exchange of content through the use of a public and a private cryptographic key pair that is obtained through a trusted authority.

156

ECM Terminology: Delivery

- **Transformation/Reformat** - Changing content from one format to the needed delivery format.
- **XML** - An established standard, based on the Standard Generalized Markup Language, designed to facilitate document construction from standard data items. Also used as a generic data exchange mechanism.

157

ECM Terminology: Preserve

- **Microfilm** (Aperture Cards, Microfiche, Microfilm Jackets, 16mm Roll Film) - (1) Fine-grain, high-resolution film used to record images reduced in size from the original. (2) Microform in the shape of a strip or roll. (3) To record microphotographs on film.
- **Optical Disc** - Primarily WORM (Write-Once, Read-Many); Optical disk on which data is recorded by the user once (and is unalterable) and can be read many times.
- **Paper** - Centuries old and, with Microfilm, one two ways to ensure that documents are readable 100 years from now, or longer.
- **SAN/NAS and CAS** (Content Addressed Storage) - Are all increasingly used for archiving content. CAS is a storage methodology designed for rapid access to fixed content.

158

Standards

159

Why Standards? (Part 1)

- Increase the rate of success in implementations
- Reduce the risk as implementations will be easier and more successful
- Increase acceptance by users and management due to the credibility that comes from systems that meet standards requirements

160

How to Avoid the Frying Pan (Part 2)

- Protect technology and product development investments
- Reduce procurement costs
- Reduce compliance costs and risks
- Improve quality
- Facilitate long term storage of electronic documents

161

Types of Standards

- De facto (DOD 5015.2)
- De facto (Any Microsoft product)
- Formalized standards organizations: American National Standards Institute (ANSI), International Standards Organization (ISO)

162

ANSI/ISO

- ARMA-Records Mgmt Standards
- AIIM-Document Mgmt Standards
- Technical Reports/ Recommended Practices
- Standards

163

Records Management-ISO-15489-1-Standard

Basic Principles and Policies:

- Characteristics of a record
- Designing and implementing records systems
- Conversion and migration
- Access, retrieval and use
- Retention & disposition

164

Records Management-ISO-15489-2-Guidelines

Implementation:

- Further explanation
- Offers a methodology
- Overview of processes and factors to consider

165

Document Management-TR-27 Electronic Imaging RFP Guidelines-Part 1

- ID fundamental steps in EIM (receipt, sorting, scanning, storing, processing
- Project Planning: Project organization, team, budget, schedule
- Process reengineering, conversion

166

Document Management-TR-27 Electronic Imaging RFP Guidelines-Part 2

- RFP Organization
- Technical Requirements
- Evaluation Tools

167

Document Management-TR-48 Integrating EDM and ERM

- EDMS/ERMS Functions
- Reference Model: Creation/Capture, Metadata Mgmt, User Mgmt, Workflow Mgmt, Reporting, Retention/Disposition, Audit trail

168

Document Management Statistical Sampling for Document Images-ISO 12032

Procedures to select and apply and apply sampling inspection plans to insure it meets your needs

- Image attributes
- Sampling plans
- Sampling size

169

Document Management ISO-15801-Trustworthiness

Ensuring acceptance of records in court:

- Policies
- Media/formats
- Procedures & Processes: Capture, indexing, file transmission, retention, destruction, system recovery, security, workflow, version control, documentation

170

Document Management: XML Wrapper

- This draft standard specifies methods for using eXtensible Markup Language (XML) for the interchange of metadata about image files, and the image files themselves.

171

Document Management: PDF-A and PDF-E

- Two different ISO draft documents responsible for specifying PDF tags for the archival storage of electronic documents.

172

For More Information

Contact ARMA or AIIM

- **AIIM Standards**

http://www.aiim.org/standards

- **ARMA Standards**

http://www.arma.org/standards/development

173

QUESTIONS?

174